DOCUMENT RESUME

ED 414 933 IR 056 791

AUTHOR	Hollas, Eric
TITLE	Technical Standards and Medieval Manuscripts.
PUB DATE	1997-04-00
NOTE	8p.; Paper presented at the Conference on Scholarly Communication and Technology (Atlanta, GA, April 24-25, 1997), see IR 056 774.
AVAILABLE FROM	Association of Research Libraries (ARL) Web site: http://www.arl.org/scomm/scat/
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Access to Information; Cataloging; Databases; Electronic Publishing; Guidelines; Higher Education; Information Sources; Information Technology; Libraries; Library Collections; *Medieval History; *Medieval Literature; Online Catalogs; Programming Languages; Scholarly Journals; Standards; Technological Advancement; User Needs (Information)
IDENTIFIERS	Electronic Resources; Manuscript Collections; MARC; *Middle Ages; SGML

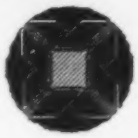
ABSTRACT

Even though medieval manuscripts represent the most voluminous surviving artifact from the Middle Ages, the very nature of this resource presents challenges for usage. In an effort to preserve medieval manuscripts and to create broader and more economical access to their contents, many libraries have in recent decades sought to provide filmed copies of their manuscripts to users. Existing and emerging electronic technologies present numerous opportunities for overcoming challenges and underscore the need to create a long-term vision for "Electronic Access to Medieval Manuscripts." Computer cataloging projects and other databases from manuscript institutions around the world represent an enormous advancement in scholarly communication in the field of manuscript studies. "Electronic Access to Medieval Manuscripts" is a three-year project to develop guidelines for cataloging medieval and renaissance manuscripts in electronic form. In addition to suggesting guidelines for content, the project will also develop standards for encoding both core-level and detailed manuscript descriptions in both MARC (Machine-Readable Cataloging) and SGML (Standardized General Markup Language). The emerging interconnectivity of MARC and SGML presents tremendous opportunities for "Electronic Access to Medieval Manuscripts." In structuring its program and goals, "Electronic Access to Medieval Manuscripts" also has sought to arrive at guidelines for encoding into MARC and SGML formats that will provide useful, economic and practical long-term alternatives to the libraries which select one of these options in the futures. (AEF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 414 933

Scholarly Communication and Technology



Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact **Richard Ekman** at the The Andrew W. Mellon Foundation.

Session #5 Technical Choices and Standards

Technical Standards and Medieval Manuscripts

Brother Eric Hollas
OSB, Director, Hill Monastic Manuscript Library
St. John's University

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Richard Ekman
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Medieval manuscripts, that is, handwritten codices produced between the fifth century and the late fifteenth century, are counted among the greatest intellectual treasures of western civilization. Manuscripts are significant to scholars of medieval culture, to art historians, calligraphers, musicologists, paleographers and other researchers for a multiplicity of reasons. They contain what remains of the classical literary corpus; and they chronicle the development of religion, history, law, philosophy, language and science from the Middle Ages into early modern times.

Even though manuscripts represent the most voluminous surviving artifact from the Middle Ages, the very nature of this resource presents challenges for usage. For one, each manuscript -- as a hand-written document -- is a unique creation. As such, copies of a particular work may contain variances that make all copies -- wherever they might be -- necessary for review by an interested scholar. Secondly, access to unique manuscripts spread across several countries or

I2560791

BEST COPY AVAILABLE

continents can be both costly and limited. A scholar wishing to consult manuscripts must often travel throughout Europe, the United States and other countries to find and study manuscripts of interest. Such research is costly and time-consuming. The universities, museums and libraries that own these manuscripts may lack the space and personnel to accommodate visiting scholars, and in some cases research appointments need to be arranged months in advance. Compounding these difficulties can be the challenge of inconvenient geography. While eminent collections reside in the great capitals of Europe, other collections of scholarly interest are housed in remote sites with no easy access at all. And finally, the uniqueness of each manuscript presents special issues of preservation. Because manuscripts represent finite and non-renewable resources, librarians concerned with the general wear and tear on manuscripts have begun to restrict access to these codices.

In an effort to preserve medieval manuscripts and to create broader and more economical access to their contents, many libraries have in recent decades sought to provide filmed copies of their manuscripts to users. This has been a long-established practice at such institutions as the British Library, the Bibliotheque National, and the Vatican Library. Additionally, some libraries have been established for the specific purpose of microfilming manuscript collections. The Institut de Recherche et d'Histoire des Textes in Paris, for example, for decades has been filming the manuscripts of the provincial libraries in France. Since its founding in 1965, the Hill Monastic Manuscript Library at Saint John's University in Minnesota has filmed libraries in Austria, Germany, Switzerland, Spain, Portugal, Malta and Ethiopia. And at the Vatican Film Library at Saint Louis University, one can find microfilms of 37,000 manuscript codices from the Biblioteca Apostolica Vaticana in Rome. Instead of traveling from country to country and from library to library, researchers may make a single trip to one of these microfilm libraries to consult texts, or, in certain circumstances, they may order microfilm copy by mail. Microfilm was a great step forward in providing access to manuscripts, and it still offers tremendous advantages of economy and democratic access to scholars. Still, there are certain limitations because in some situations researchers must visit the microfilm institutions to consult directly, and the purchase of microfilm -- even if ordered from a distance -- can entail long waits for delivery. And compounding these difficulties can be the inconsistency or inadequacy of existing descriptions of medieval manuscripts.

Access to manuscripts in particular collections is guided by the finding aids that have been developed through the centuries. The medieval shelf list has given way to the modern catalogue in most cases, but challenges in locating particular manuscripts and in acquiring consistent information abound. Traditionally, libraries in Europe, the United States, and elsewhere have published manuscript catalogues to describe their handwritten books. These catalogues are themselves scholarly works that combine identification of texts with a description of the codex as a physical object. Although these catalogues are tremendously valuable to scholars, they are not without their shortcomings. With respect to manuscript catalogues, there is presently no agreement within the medieval community on the amount and choice of detail reported, on the amount of scholarly discussion provided and on the format of presentation. Moreover, to consult these published books in the aggregate requires access to a research library prepared to maintain an increasingly large collection of expensive and specialized books. And beyond that, the production of a modern catalogue requires expertise of high caliber and the financial resources that facilitate the work. Because many libraries do not have such resources available, many collections have gone uncatalogued or have been catalogued only in an incomplete fashion. The result for the scholar is a paucity of the kind of information that makes manuscript identification and location possible.

Existing and emerging electronic technologies present extraordinary opportunities for overcoming these challenges and underscore the need to create a long-term vision for *Electronic Access to Medieval Manuscripts*. Electronic access both to manuscript images as well as to bibliographic information presents remarkable opportunities. For one, the distance between the manuscript and the reader vanishes -- providing the opportunity for a researcher anywhere to consult the image of a manuscript in even the remotest location. Secondly, electronic access obviates the security issues and the preservation concerns that accompany usage. Furthermore, electronic access will permit the scholar to unite the parts of a manuscript that may have been taken apart, scattered, and subsequently housed at different sites. It also allows for image enhancement and manipulation that conventional reproductions simply do not make available. Electronic access will also make possible comprehensive searches of catalogue records, research information, texts and tools -- with profound implications in terms of cost to the researcher and a more democratic availability of materials to a wider public.

One may imagine a research scenario that contrasts sharply with the conventional methods that have been the mainstay of manuscript researchers. Using a personal computer in an office, home, educational institution or library, scholars will be able to log on to a bibliographic utility (i.e. RLIN or OCLC) or on to an SGML database on the World Wide Web and browse catalogue records from the major manuscript collections around the world. To make this vision a reality requires adherence to standards, however -- content standards to insure that records include the information that scholars need, and encoding standards to insure that that information will be widely accessible both now and in the future.

This point may be demonstrated by considering several computer cataloguing projects developed since the mid-1980's. These efforts include the Benjamin Catalogue for the History of Science, the International Computer Catalog of Medieval Scientific Manuscripts in Munich, the Zentralinventar Mittelalterlicher Handschriften (ZIH) at the Deutsche Staatsbibliothek in Berlin, MEDIUM at the Institut de Recherche et d'Histoire des Textes in Paris and PhiloBiblon at the University of California, Berkeley. The Hill Monastic Manuscript Library has also embarked on several electronic projects to increase and enhance scholarly access to its manuscript resources. In 1985, Thomas Amos, then Cataloguer of Western Manuscripts at HMML, began development of the *Computer Assisted Cataloguing Project*, a relational database which he used to catalogue manuscripts from Portuguese libraries filmed by HMML.

These electronic databases as well as others from manuscript institutions around the world represent an enormous advancement in scholarly communication in the field of manuscript studies. As in the case of printed catalogues and finding aids, however, these data management systems fall short of the ideal on several counts. First, each is a local system that must be consulted on site or purchased independently. Second, the development and maintenance of these various databases involve duplication of time, money and human resources. All rely on locally-developed or proprietary software, and this has posed problems for the long-term maintenance and accessibility of the information. Finally, and probably most importantly, each system contains its own unique set of data elements and rules and procedures for data entry and retrieval. When each of these projects was begun, its founders decided independently what information about a manuscript to record, how to encode it and how to retrieve it. Each of the databases adopted a different solution to the basic problems of description and indexing, and the projects differed from each other with regard to completeness of the data entered and the modes in which it could be retrieved.

The lessons to be drawn from these experiences are clear and enunciate the hazzards for the

future if distinctively different approaches are not pursued. First of all, local institutions could not maintain locally developed software and systems. In the instances of projects that chose to rely on proprietary software, it became apparent that the latter was dependent on support from the manufacturer, whose own longevity in business could not be guaranteed, or who could easily abandon such software programs when advances provided new opportunities. Furthermore, experience has demonstrated that it is not always easy to translate such material into other formats, and if modified it poses the same problems of maintenance as locally developed software. Beyond that, different projects made substantially different decisions about record content, and those decisions were sometimes influenced by the software that was available. This lack of consistency made it difficult to disseminate the information gathered by each project, and for their part funding agencies were reluctant to continue their support for such limited projects. All of which reiterates the fundamental need for content standards to insure that records include the information that scholars need and encoding standards to insure the wide accessibility of that information both now and into the future. It is the objective of *Electronic Access to Medieval Manuscripts* to address these issues.

Electronic Access to Medieval Manuscripts is sponsored by the Hill Monastic Manuscript Library, Saint John's University, Collegeville, Minnesota, in association with the Vatican Film Library, Saint Louis University, and has been funded by a grant from The Andrew W. Mellon Foundation. It is a three-year project to develop guidelines for cataloguing medieval and renaissance manuscripts in electronic form. For this purpose it has assembled an international team of experts in manuscript studies and library and information science which will examine the best current manuscript cataloging practice in order to identify the information appropriate to describing and indexing manuscripts on two levels, core and detailed. Core level descriptions, which will contain the basic or minimum elements required for the identification of a manuscript, will be useful for describing manuscripts that have not yet been fully cataloged, and may also be used to give access to detailed descriptions, or to identify the sources of digital images or other information extracted from manuscripts. Guidelines for detailed or full descriptions will be designed to accommodate the kinds of information found in full scholarly manuscript cataloging.

In addition to suggesting guidelines for content, *Electronic Access to Medieval Manuscripts* will also develop standards for encoding both core-level and detailed manuscript descriptions in both MARC and SGML. The MARC (Machine-Readable Cataloging) format underlies most electronic library catalogs in North America and the United Kingdom, and it is used also as a vehicle for international exchange of bibliographic information. MARC bibliographic records are widely accessible through local and national databases, and libraries with MARC-based cataloguing systems can be expected to maintain them for the foreseeable future. SGML (Standardized General Markup Language) is a platform-independent and extremely flexible way of encoding electronic texts for transmission and indexing. It supports the linking of texts and images, and SGML-encoded descriptions are easily converted to HTML for display on the World Wide Web. In developing standards for SGML encoding of manuscript descriptions, *Electronic Access to Medieval Manuscripts* will work closely with the *Digital Scriptorium*, a project sponsored jointly by the Bancroft Library at the University of California, Berkeley, and the Butler Library at Columbia University.

The project working group for *Electronic Access to Medieval Manuscripts* consists of representatives from a number of North American and European institutions. Drafts produced by the working group will be advertised and circulated to the international community of manuscript scholars for review and suggestions. The cataloguing and encoding guidelines that result from the work of the project will be made freely available to any institution that wishes to

use them.

For the purposes of *Electronic Access to Medieval Manuscripts*, the standards for cataloguing medieval manuscripts are crucial, but so too is the application of content standards to the two encoding standards whose existence and ubiquitous usage address the issues noted earlier. At the risk of stating the obvious, *Electronic Access to Medieval Manuscripts* has chosen to work with two existing and widely used encoding standards because it is unwise for medievalists to reinvent the wheel and waste resources on solutions that are temporary and which will require added resources to take them into future applications.

With regard to encoding standards, the universal acceptance of MARC and the accessibility of MARC records on-line make it a particularly attractive option. But there are other compelling reasons that make MARC an excellent choice. First, most libraries already have access to a bibliographic utility (such as OCLC and RLIN) that utilizes MARC-based records, and these institutions have invested considerable resources in creating catalogue records for their printed books and other collections. Second, since most catalogue records for printed books and reference materials are already in MARC-based systems, placing manuscript records in the same system makes good sense from the standpoint of proximity and one-stop searching. Third, by using MARC, local libraries need not develop or maintain their own database systems. Finally, although it may be unrealistic to expect that all manuscript catalogue records will one day reside in a single database, therefore allowing for a universal search of manuscript records, it is far more likely that a majority of manuscript institutions in the United States will be willing to place their manuscript records in this bibliographic utility rather than in other existing environments. Thus the value of selecting MARC as an encoding standard seems clear. MARC systems exist; they are widely accessible; they are supported by other broader interests; and enough bibliographic data already exists in MARC to guarantee its maintenance or its automatic transfer to any future platform. In USMARC (RLIN and OCLC databases) there are already a significant number of records for medieval manuscripts or microfilms of them, prepared and entered by the various institutions that hold these items. Regrettably, there is generally little consistency in description, indexing or retrieval for these records; all of which points back to the need for standards for content as well as encoding standards. Furthermore, MARC as it currently exists has limits in its abilities to describe medieval manuscripts (e.g.: it does not provide for the inclusion of *incipits*), but nonetheless it offers possibilities for short records that point to broader sets of data in other contexts. Still, MARC, with its records in existing bibliographic databases, is particularly advantageous for small institutions with few manuscript holdings, and it remains for them perhaps the most promising vehicle for disseminating information about their collections.

The second viable encoding option, particularly in light of the recent success of the Archival Finding Aid Project at the University of California, Berkeley, is the use of Standard Generalized Markup Language (SGML). As a universal standard for encoding text, SGML can be used to encode and index catalogue records and other data including text, graphics, images and multimedia objects such as video and sound. A more flexible tool than MARC, SGML is more easily adapted to complex hierarchical structures such as traditional descriptions of medieval manuscripts, and it offers broad possibilities for encoding and indexing existing, as well as new, manuscript catalogues. As an encoding scheme, SGML demonstrates its value as a non-proprietary standard. In many respects it is much more flexible than MARC or any established database program, and it is possible to write a Document Type Definition (DTD) taking into account the particular characteristics of any class of document. SGML offers the further advantage that encoded descriptions can be linked directly to digital images, sound clips

(e.g., for musical performances) or other bodies of digital information relating to a manuscript. Numerous initiatives using SGML suggest great promise for the future. The experience of the American archival profession with the Encoded Archival Description (EAD) suggests that this can be a good approach to encoding manuscript descriptions, which have many structural analogies to archival finding aids. The Canterbury Tales project, based at Oxford, has demonstrated that SGML, based on a Text Encoding Initiative (TEI) format, can be used successfully to give sophisticated access to images of manuscripts, text transcriptions and related materials. In addition, several English libraries have already experimented with SGML DTD's, mostly TEI-conformant, for manuscripts. And finally, MASTER, an Oxford-based group, is interested in developing a standard DTD for catalogue descriptions of medieval manuscripts, and it and *Electronic Access to Medieval Manuscripts* have begun to coordinate their efforts toward achieving this common goal.

The emerging interconnectivity of MARC and SGML presents tremendous opportunities for *Electronic Access to Medieval Manuscripts*. Currently there is work on a DTD for the MARC format that will allow automatic conversion of MARC encoded records into SGML. Recently, a new field (856) was added to the MARC record that will accommodate web addresses. Implementation of this field will allow researchers seeking access to a cataloguing record in a bibliographic utility to read the URL (Universal Resource Locator) and then enter the address into a web browser and link directly to a website containing a detailed manuscript record or other scholarly information. In the future, for researchers who enter the bibliographic utility through a web browser, this will be an active hypertext link. *Electronic Access to Medieval Manuscripts* envisions an environment in which institutions can enter their manuscript catalogue records into MARC, display them in a bibliographic utility to maximize economy and access, and then embed a hypertext link to a more detailed catalogue record, an image file or scholarly information on an SGML server.

It has been the cumulative experience of recent years that has shaped the development and goals of *Electronic Access to Medieval Manuscripts*. Concerned with arriving at standards for cataloguing manuscripts in an electronic environment, the project seeks to provide standards for both core and full or detailed level manuscript records that will serve the expectations and needs of scholars who seek consistent information from one library to another, while they will afford flexibility to those cataloguers and libraries wishing to provide various levels of information about their individual manuscripts. In structuring its program and goals, *Electronic Access to Medieval Manuscripts* also has sought to arrive at guidelines for encoding into MARC and SGML formats that will provide useful, economic and practical long-term alternatives to the libraries which select one of these options in the future.

. . . . . .

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC
Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: August 1, 1997